



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

10/621,474

07/17/2003

Hiroshi Nomiya

JP920020109US1

2462

25259

7590

06/22/2006

IBM CORPORATION  
3039 CORNWALLIS RD.  
DEPT. T81 / B503, PO BOX 12195  
RESEARCH TRIANGLE PARK, NC 27709

EXAMINER

ROSE, HELENE ROBERTA

ART UNIT

PAPER NUMBER

2163

DATE MAILED: 06/22/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b> 10/621,474	<b>Applicant(s)</b> NOMIYAMA ET AL.	
	<b>Examiner</b> Helene Rose	<b>Art Unit</b> 2163	

**-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --**

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 4/4/06.
- 2a) ☒ This action is **FINAL**.                      2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 8,10-13 and 15-20 is/are pending in the application.  
     4a) Of the above claim(s) 1-7,9,14 is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 8,10-13 and 15-20 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
     Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
     Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
     a) ☐ All    b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)  | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)   | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)             |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date <u>5/26/05</u> | 6) <input type="checkbox"/> Other: _____  |

**Detailed Action**

1. In response to communications filed on April 04, 2006, claims 8 and 10-13 have been amended; claims 1-7,9 and 14 have been cancelled and new claims 15-20 have been added per applicant request. Therefore, claims 8, 10-13, and 15-20 are presently pending in this application.
2. Applicant's arguments filed on April 04, 2006 have been fully considered have been fully considered (MPEP 714.04; 37 CFR 1.111) but they are not persuasive.

**Information Disclosure Statement**

3. The information disclosure statement filed 5/26/05 fails to comply with 37 CFR 1.98(a)(2), in which the cited JP foreign patent documents 2001-060165; 2001-325272; 07-006076 as well as the non-patent literature documents cited as other art; requires a legible copy of each cited foreign patent document; each non-patent literature publication or that portion which caused it to be listed; and all other information or that portion which caused it to be listed. Applicant is required to indicate on the Information Disclosure Form what is to be considered whether it's the abstract or full document as it relates to the foreign patents, neither was specified nor cited on the form, and a full translation of the non-patent literatures must be submitted. It has been placed in the application file, but the information referred to therein has not been considered.

**Claim Rejections – 35 U.S.C 112**

5. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

6. Claim 13 is rejected under 112, second paragraph. Claim 13 recite the following limitation "or", this limitation renders the claim vague and indefinite, because the term "or" is considered to be alternative language. Therefore, the limitation renders the claim vague and indefinite, because it is unclear as to how the examiner should interpret the claim limitation as it relates to "or".

**Claim Rejections – 35 U.S.C 103**

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

6. Claims 8,10-13, and15-20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Ishikawa et al (US Patent No. 5,848,407, Date of Patent: December 8, 1998) in view of Gutierrez et al (US Publication No. 2003/0046276, Filing Date: September 06, 2001) and further in view of Finseth et al (US Patent No. 6,271,840, Date of Patent: August 7, 2001).

**(CURRENTLY AMENDED) Claims 8 and 10:**

Regarding claims 8 and 10, Ishikawa teaches an information search method for crawling a web site via a network using a computer, said method comprising the steps of:

acquiring a web page as initial information and storing source code into a storage device (see Figure 4, all features, Ishikawa);

reading the source code of said web page from said storage device (Figure 4(see Figure 4, wherein text body is taking into account to be the source code prescribed in webpage, Ishikawa);

conducting a structure analysis of said web page (see Figure 4, wherein parent document list is conducting a structure analysis of web page, Ishikawa) and column 19, lines 4-7, wherein user can realize document by reading the summary of each particular hypertext document without calling each particular hypertext document, Ishikawa), **wherein the structure analysis includes the steps of:**

Ishikawa discloses all the limitations above. However, Ishikawa is silent with respect wherein the source code is HTML. On the other hand, Gutierrez discloses **reading an HTML document of a web page as an analyzing object** (Figure 5, diagram 505, Gutierrez). A skilled artisan would have

been motivated to combine as suggested by Gutierrez, by clearly defining as well as stating that the source code is HTML, for creating hypertext documents for the World Wide Web, easy editing, and viewable documents.

conducting a temporary block analysis based on a description of HTML tags of the HTML document (page 3, section [0036], wherein the resulting data is often initially stored in temporary database storage area, Gutierrez),

using the HTML tags to temporarily divide the HTML document into blocks (column 19, lines 55-57, wherein a composition or an article is divided into a number of portions, and each portion of the composition is written in on of the hypertext documents, Ishikawa), and

identifying unnecessary information elements in the HTML document (Figure 6, diagram 630 and further defined on page 5, section [0052], wherein the matched data is ordered by weighting information included in search index so that more relevant data is more likely to be displayed before less relevant data, wherein the results are formatted using formatting language such as HTML, Gutierrez).

Ishikawa in view of Gutierrez discloses all the limitations above. However, Ishikawa in view of Gutierrez do not disclose wherein the unnecessary information elements are plural information elements that include an OBJECT IMAGE having a same Uniform Resource Locator (URL); nor do they disclose wherein the OBJECT IMAGE describes a type of media used to display the HTML document reading an HTML document of a web page as an analyzing object. On the other hand, Finseth discloses wherein the unnecessary information elements are plural information elements that include an OBJECT IMAGE having a same Uniform Resource Locator (URL) (see abstract, wherein search engine results or a list of URL's are passed to a web crawler that retrieves the web page and other media information present at the associated URL, wherein defined in applicants specification within the PG Publication, paragraph [0059], wherein OBJECT\_IMAGE, represents all media types, Finseth), and wherein the OBJECT IMAGE describes a type of media used to display the HTML document (column 5, lines 32-35 and lines 62-67, wherein rendering engines may include web document or HTML, Finseth).

A skilled artisan would have been motivated to combine as suggested by Ishikawa in view of Gutierrez, removing duplicate information to provide the a user with relevancy and enabling faster perusal of search engine output as it relates to the world wide web.

storing a result of the analysis into said storing device (column 6, lines 31-35, Ishikawa);

calculating a degree of significance of a web site linking from said web page, based on the result of said structure analysis stored in said storage device (see Figure 5, all features and column 10, lines 18-25, Ishikawa); and

accessing the web site depending on the calculated degree of significance to acquire contents thereof (column 3, lines 4-19, Ishikawa), and storing them into said storage device (see Figure 3, diagram 8, Ishikawa).

Claim 11:

Regarding claim 11, the combination of Ishikawa in view of Gutierrez and further view in Finseth teaches wherein said structure analyzing means associates mutually relevant information elements with each other, among information elements contained in said source code (see Figure 4, wherein word list is the relevant information elements and text body is the information elements contained in source code, and source code is considered to be taking into account the prescribed web page, and column 9, lines 4-16, Ishikawa).

Claim 12:

Regarding claim 12, the combination of Ishikawa in view of Gutierrez and further view in Finseth teaches wherein said significance calculating means selects plural strategies as strategies for calculating the degree of significance of (column 10, lines 50-61, wherein selecting a plurality of candidates for a key word, Ishikawa) said web site (column 10, lines 41-46, wherein the user calls the HTML document in the world wide web, Ishikawa), and uses them by giving weights thereto, respectively (column 11, lines 47-65, wherein weights are assigned he number of keywords is two or more, it is applicable that an estimated value for one particular hypertext document be set to a value N times (N is two or more) as high as a sum of the products  $TF \cdot IDF$  calculated for all keywords when N particular words agreeing with

N keywords appear in the particular hypertext document and . where two particular words agreeing with two keywords are used in one particular hypertext document close to each other within 20 characters, it is applicable that an estimated value for the unified particular hypertext document be doubled, Ishikawa).

(CURRENTLY AMENDED) Claim 13:

Regarding claim 13, Ishikawa view of Gutierrez and further in view of Finseth teaches a program product for controlling a computer so as to analyze an HTML document structure, said program causing said computer to execute:

a first process of reading an HTML document being a processing object from a memory (column 19, lines 4-7, Ishikawa), blocking information elements forming said HTML document based on tags of said HTML document (column 24, lines 39-48, wherein blocking means to block relevant information elements among information elements included in HTML document according to applicants specification, Ishikawa), and storing blocked structural data of said HTML document into the memory (column 24, lines 52-57, Ishikawa); and

a second process of reading the blocked structural data of said HTML document from said memory, updating block structures of said HTML document by associating the information elements that are mutually relevant in terms of a meaning (column 17, lines 25-30, wherein a revised occurrence frequency TF for the particular hypertext document for each of the particular hypertext documents, Ishikawa), and storing the updated structural data into the memory (see Figure large volume of hypertext documents stored in the hypertext document managing unit 8 and column 17, lines 17-19, Ishikawa), wherein said second process includes the step of:

identifying an unnecessary information element in term of a purpose of document structure analysis, wherein an information element is deemed to be unnecessary if the information element includes an OBJECT IMAGE that includes a Uniform Resource Locator (URL) that has been used by another information element in the HTML document, wherein the OBJECT IMAGE describes a type of media used to describe the HTML document (REFER to claim 8, wherein these limitation have already been addressed, Gutierrez); and

**merging said information elements or dividing a block based on contents of said information elements** (REFER to claim 8, wherein the limitation for dividing a block based on said contents, Ishikawa); **and**

**merging the block structure based on information contained in each block** (column 16, lines 52-67, wherein second embodiment and third embodiment is combined, and wherein the third group of documents D81, D83, D86 set to the 19<sup>th</sup> rank is reset to the forth rank and a combined group of the particular hypertext documents D83, D85 and D86 and the parent documents D80 and D81 reset to the forth rank as displayed in Figure 9, all features, Ishikawa).

**(NEW) Claim 15:**

Regarding claim 15, Ishikawa in view of Gutierrez and further in view of Finseth teaches reading an HTML document of a web page as an analyzing object (**REFER** to claim 8, wherein this limitation has already been addressed);

conducting a temporary block analysis based on a description of HTML tags of the HTML document (**REFER** to claim 8, wherein this limitation has already been addressed, Gutierrez);

using the HTML tags to temporarily divide the HTML document into blocks (**REFER** to claim 8, wherein this limitation has already been addressed, Ishikawa);

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT\_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT\_IMAGE describes a type of media used to display the HTML document (**REFER** to claim 8, wherein this limitation has already been addressed, Gutierrez); and

deleting any block in the HTML document that is deemed to be structurally meaningless, wherein a block is deemed to be structurally meaningless if that block has only unnecessary information elements (columns 8-9, lines 65-67 and line 1-6, wherein removing from a plurality of character strings existing in a body of collected reference document to form a text body, Ishikawa);



merging relevant information elements in a same block into one composite element (REFER to claim 13, wherein this limitation has already been addressed, Ishikawa).

(NEW) Claim 16:

Regarding claim 16, Ishikawa in view of Gutierrez and further in view of Finseth teaches wherein the unnecessary information elements include OBJECT\_ANCHORS that have a same title, wherein an OBJECT\_ANCHOR describes a correlation between the HTML document and elements in another web page (column 9, lines 52-64 and column 10, lines 48-63, Finseth).

(NEW) Claim 17:

Regarding claim 17, Ishikawa in view of Gutierrez and further in view of Finseth teaches wherein the unnecessary information elements include OBJECT\_TEXT\_BLOCKS that have the same description of text in a block (Figures 6 and 7, all features, further define in column 9, lines 52-55, wherein an image map is associated with the rendered image then clicking upon different areas of the rendered image will have the same results as those set forth in the description of Figure 6, Finseth).

(NEW) Claim 18:

Regarding claim 18, Ishikawa in view of Gutierrez and further in view of Finseth teaches wherein the relevant information elements that are merged are from a group that includes the OBJECT\_IMAGE, OBJECT\_ANCHOR and OBJECT\_TEXT\_BLOCKS (Figure 4, all features, Ishikawa).

(NEW) Claim 19:

Regarding claim 19, Ishikawa in view of Gutierrez and further in view of Finseth teaches wherein a method for eliminating ambiguity of a specified topic being searched during a web crawling, the method comprising:

presenting relevant keywords to a user during web crawling, wherein the relevant keywords describe multiple attributes of a term that has an ambiguous meaning, and wherein the user is afforded

an ability to specify keywords that have a minus degree of significance to a meaning intended by the user for web crawling (column 1, lines 53-67, Ishikawa, wherein a plurality of identification titles indicating the retrieval documents are arranged in increasing order, wherein increasing is equivalent to minus, when the user selects the identification titles displayed on the unit one after another in arranged order the retrieval document indicated by the selected identification title is read out from the document, wherein the title is equivalent to a keyword, Ishikawa); and

narrowing down crawling objects by eliminating user-specified keywords that have a minus degree of significance, thereby eliminating ambiguity of a term being searched (columns 8-9, lines 65-67 and lines 1-6 and lines 13-16, wherein removing is equivalent to deleting, from a plurality of strings existing in a body of collected reference document to form a text body, and the text body is written in the document information entry space, and wherein a plurality of words used in the text body the title and the anchor sentence are written in the document, wherein words is equivalent to term, Ishikawa).

(NEW) Claim 20:

Regarding claim 20, Ishikawa in view of Gutierrez and further in view of Finseth teaches a web crawler comprising:

an initial site acquiring section, wherein the initial site acquiring section specifies a Uniform Resource Locator (URL) of a home page of a specific web site from which information is to be collected, (Figure 5, diagram 505, also see page 4, section [0047], Gutierrez) and wherein initial web sites to be searched are obtained through the use of keywords in a search engine (Figure 5, diagram 510, also see page 4, section [0047], also see page 5, section [0050], Gutierrez), and wherein the initial web sites that are initially set for web crawling (Figure 5, diagrams 595 and 515, Gutierrez);

a document structure analysis section for performing document structure analysis for a web page of initial sites, wherein the document structure analysis includes the step of:

reading an HTML document of a web page as an analyzing object REFER to claim 8, wherein this limitation has already been addressed);

conducting a temporary block analysis based on a description of HTML tags of the HTML document (REFER to claim 8, wherein this limitation has already been addressed);

using the HTML tags to temporarily divide the HTML document into blocks (REFER to claim 8, wherein this limitation has already been addressed);

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT\_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT\_IMAGE describes a type of media used to display the HTML document (REFER to claim 8, wherein this limitation has already been addressed); and

a significance calculating section for calculating degrees of significance of web sites that are acquired by web crawling, wherein the degrees of significance are based on a result of the document structure analysis performed by the document structure analysis section (column 4, lines 22-39 and column 10, lines 18-26, Ishikawa), and wherein calculating degrees of significance extends a Fish-search crawling technique by basing the calculating on strategies specified by a user and information elements added to anchors through the document structure analysis section, and wherein objects of crawling are dynamically determined depending on the degree of significance (column 17, lines 51-62, Ishikawa); and

a crawling executing section for executing a process of acquiring web sites by crawling based on the results of the degrees of significance calculated by the significance calculating section (column 4, lines 35-40, Ishikawa).

### **Response to Arguments**

7. Applicant argues the prior art fails to teach “conducting a structure analysis of said web page, wherein the structure analysis includes the steps of reading an HTML document of a web page as an analyzing object, conducting a temporary block analysis based on a description of HTML tags of the HTML document as it relates to claim 8”

Applicant argues the amended claim limitation. In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., *reading an HTML document of a web page as an analyzing object, conducting a temporary block analysis based on a description of HTML tags of the HTML document*) are not recited in

the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

8. In reference to claims 10, 13, and 15-20, applicant continues to argue the amended claims and the newly added claims. It is noted that the amended claims and newly added claims in which the applicant relies on fails to show within in the original claim language. Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

9. Therefore, Applicant's arguments filed on April 04, 2006, with respect to the rejected claims in view of the cited references have been considered but are moot in view of applicant's amended claims necessitate new ground(s) of rejection.

#### ***Conclusion***

10. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

11. A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

**Prior Art of Record**

(The prior art made of record and not relied upon is considered pertinent to applicant's disclosure)

1. Ishikawa et al (US Patent No. 5,848,407) discloses a hypertext document and anchor sentences of parent documents for the hypertext document are registered with an hypertext document identifier as document information for each of hypertext documents having reference relationships with each other.
2. Finseth et al (US Patent No. 6,271,840) discloses a visual index method provides graphical output from search engine results or other URL lists.
3. Gutierrez et al (US PG Publication No. 2003/0046276) discloses a system and method for searching a database from a computer network, wherein a client computer sends a search request to a search engine and search engine prepares a database request, and wherein the search engine sends the database request to one or more servers that include database management systems, such as IBM's DB2.TM; the servers receive the request and extract responsive data from the databases being managed by the database management system, wherein the extracted data is returned to the search engine which is then formatted and returned to the client.

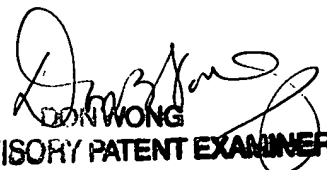
**Point of Contact**

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Helene R. Rose whose telephone number is (571) 272-0749. The examiner can normally be reached on 8:00am - 4:30pm M-F.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Don Wong can be reached on (571) 272-1834. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Helene R Rose  
Technology Center 2100  
June 6, 2006

  
DON WONG  
**SUPERVISORY PATENT EXAMINER**